

ケプストラムに基づく声道断面積関数の推定

石 田 義 久
鎌 田 弘 之*
小 川 康 男

Estimation of Vocal Tract Area Functions Based on Cepstrum

by

Yoshihisa ISHIDA, Hiroyuki KAMATA and Yasuo OGAWA

Synopsis. This paper presents a simple method for the estimation of vocal tract area functions.

This method is summarized as follows. First, the frequency spectra of speech waves are detected by the filter bank which consists of 17 band-pass filters. Utilizing inverse DFT to this spectra, one can obtain the impulse response of a vocal tract. Secondly, the step-down method is applied to the impulse response, and vocal tract area functions are estimated.

1. ま え が き

マイクロコンピュータ技術の発達によって、数年前のミニコンに匹敵しうる機能を備えたパーソナルコンピュータが実現しつつある。又、日本語処理機能も充実し、マイコン本来の適用分野である FA (ファクトリオートメーション) のみならず、OA (オフィスオートメーション) 用としても積極的に利用されるようになり、ますますその応用範囲が拡大してきている。

筆者らも、これまでパーソナルコンピュータのもつ汎用性、取扱い易さなどの特徴に着目し、主に聴覚障害者を対象とした特殊教育機器への応用を進めてきた。昭和59年6月に川崎市立ろう学校に寄贈されたスピーチトレーナー7号機は¹⁾、Z-80を搭載したパーソナルコンピュータによるもので、従来のトレーナーに対して次のような改善が図られている。

- (1) 音声の基本的な情報要素であるピッチパタンやスペクトルパタンをカラーで視覚化し、手本パタンと被訓練者による練習パタンとの差異を明らかにした。
- (2) 制御プログラムの主要な部分に BASIC 言語を導入し、対話形式で機械を操作できるようにした。
- (3) 従来のハードウェアを主体とした回路構成を、ソフトウェアによるプログラム制御方式

* 明治大学大学院工学研究科

に改め、ハードウェア部の簡素化を図った。

さて、これまでのスピーチトレーナーは、主として小学校低学年から高校生程度までを対象としていた。このため、発話訓練に最も重要な時期にある就学以前の幼児には、音声情報の表現方法が難しく、理解させるのに若干の時間を要した。これを解決するには、例えばイラストなどを導入して、幼児に理解しやすい形で音声情報を表示する必要がある。特に、発声音の音韻性を表現するために用いているスペクトルグラフは、音声の分析的特徴を与えるのに便利であるが、発話の際の運動感覚を直観的に理解させることが困難である。

そこで、本研究では、調音方法を直観的に理解させる一つの試みとして、声道形状を簡単に求める方法を開発し、更にこれをほぼ実時間でディスプレイ上に表示する装置を試作した。この方法は、スペクトルグラフの複素ケプストラムを利用するもので、帯域フィルタ群によって得られたスペクトルに対して、対数変換、逆フーリエ変換などの数学的处理を施すことによって、比較的高い精度で声道形状を求めることができる。

以下、パーソナルコンピュータを利用して、複素ケプストラムにより声道形状を推定する方法について述べる。

2. 複素ケプストラムによる声道断面積関数の推定

2.1 原 理

音声波形を音源によって励振された声道の出力であると考え、声道の特性を線形時不変と仮定すれば、音声波形は予測可能であり、 n 時点における予測値 S_n は、過去のあいつぐ p 個のサンプル値 ($S_{n-1}, S_{n-2}, \dots, S_{n-p}$) から、次式によって推定できる²⁾。

$$\hat{S}_n = -(a_1 S_{n-1} + a_2 S_{n-2} + \dots + a_p S_{n-p}) \dots\dots\dots (1)$$

ここで、 a_i ($i=1 \sim p$) は線形予測係数であり、負の符号は、後に述べる音声の生成モデルの形を整えるためである。

実際の音声信号値 S_n との予測誤差 ε_n は

$$\varepsilon_n = S_n - \hat{S}_n = S_n + (a_1 S_{n-1} + a_2 S_{n-2} + \dots + a_p S_{n-p}) = \sum_{i=0}^p a_i S_{n-i}, \quad a_0 = 1 \dots\dots\dots (2)$$

(2)式で、予測誤差 ε_n を入力とし、音声波形を出力とするようなフィルタの入出力関係を見ると、フィルタの伝達関数 $H(z)$ は、

$$H(z) = \frac{S(z)}{E(z)} = \frac{1}{1 + \sum_{i=1}^p a_i z^{-i}} \dots\dots\dots (3)$$

与えられる。ただし、 $S(z)$ 、 $E(z)$ は、それぞれ S_n 、 ε_n の z 変換である。

予測誤差は、一般に複雑な波形であるが、有声音区間では準周期的であり、無声音区間では不規則である。したがって、予測誤差を入力とする(3)式のフィルタは、音声生成モデルの一つと考えられ、声道の特性を表わすことになる。実際には、鼻音のように、鼻腔の共振を考慮して、(3)式に零点を加えた、いわゆる極零モデルを考える必要があるが、次数 p が十分高ければ、

(3)式のような全極形モデルでも音声波形のよい近似を与える。

さて、声道の周波数伝達特性は、 $z=e^{j\omega}$ において

$$|H(e^{j\omega})|=1/|1+\sum_{i=1}^p a_i e^{-i j\omega}| \dots\dots\dots (4)$$

となる。一方、音声波形の対数スペクトルは、(3)式より

$$\log|S(e^{j\omega})|=\log|H(e^{j\omega})|+\log|E(e^{j\omega})| \dots\dots\dots (5)$$

で与えられる。ここで、上式から音源のスペクトル成分を分離することができれば、

$$\log|S(e^{j\omega})|=\log|H(e^{j\omega})| \dots\dots\dots (6)$$

となり、音声波形のスペクトル成分から声道の伝達特性を推定できることになる。音源のスペクトルは、ほぼ一定の特性（有声音の場合、近似的に -12dB/oct の包絡特性、無声音の場合は、ほぼ平坦な包絡特性）をもっており、プリエンファシスのような簡単な前処理によって、これを除去することができる。そこで、

$$\log|S(z)|=\log|H(z)|=-\log\left|\sum_{i=0}^p a_i z^{-i}\right|=-\log|A(z)|$$

$$\text{ただし、} A(z)=\sum_{i=0}^p a_i z^{-i}$$

が成り立つとすれば

$$\hat{A}(z)\triangleq\log[A(z)]=\log|A(z)|+j\arg[A(z)]=-\log|S(z)|+j\arg[A(z)] \dots\dots\dots (7)$$

である。 $\hat{A}(z)$ がある数列 $a(n)$ の z 変換と考える場合、 $a(n)$ が実数の因果性で安定な数列（最小位相数列）であれば、 $\hat{A}(z)$ の実数部 $(-\log|S(z)|)$ から $a(n)$ が完全に決定される³⁾。この $a(n)$ のことを複素ケプストラムと呼ぶ。

$a(n)$ のフーリエ変換 $\hat{A}(e^{j\omega})$ は、(7)式において $z=e^{j\omega}$ とおけば

$$\hat{A}(e^{j\omega})=-\log|S(e^{j\omega})|+j\arg[A(e^{j\omega})]$$

となり、 $-\log|S(e^{j\omega})|$ の逆フーリエ変換が $a(n)$ の偶関数部（これを $C(n)$ と表わす）に等しいことから

$$C(n)=\frac{a(n)+a(-n)}{2}$$

である。最小位相数列に対しては、因果性

$$a(n)=0, \quad n<0$$

より

$$a(n)=C(n)\cdot U_+(n) \dots\dots\dots (8)$$

ここで

$$U_+(n)=\begin{cases} 0, & n<0 \\ 1, & n=0 \\ 2, & n>0 \end{cases}$$

となる。

音声波形のスペクトル $S(e^{j\omega})$ から複素ケプストラム $a(n)$ が求まれば、次の再帰的計算によ

って、音声生成モデルのフィルタ係数 a_i を求めることができる。

$$ia_i = -\sum_{n=0}^{i-1} (i-n) \hat{a}(i-n) a_n \quad (a_0=1) \dots\dots\dots (9)$$

このように、対数音声スペクトル $\log|S(e^{j\omega})|$ から、声道の伝達関数の推定値であるような逆フィルタ $1/A(z)$ が求まる。

一方、声道を断面積の異なる p 個の音響管が接続されているものと仮定して、図 1 (a) に示す声道断面積モデルを考える。いま、接続点 i における反射係数を k_i とすれば

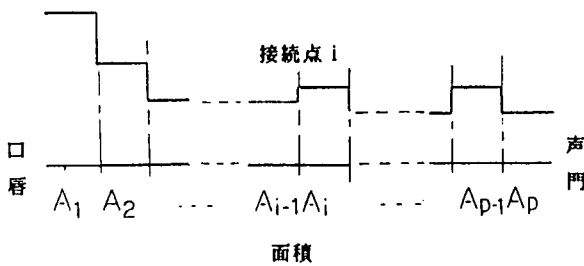
$$k_i = \frac{A_{i-1} - A_i}{A_{i-1} + A_i}, \quad i=1, 2, \dots, p$$

または

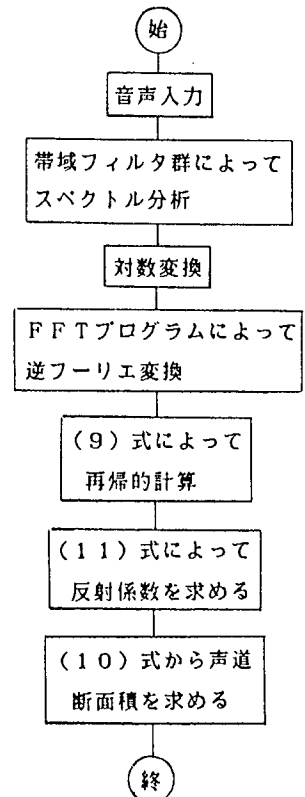
$$A_{i-1} = \frac{1+k_i}{1-k_i} A_i \dots\dots\dots (10)$$

と表わすことができる。この反射係数 $\{k_i\}$ は前述の逆フィルタ $1/A(z)$ が与えられたなら、次のステップダウン過程から再帰的に求めることができる⁴⁾。

$$a_{i-1,j} = \frac{a_{ij} - k_i a_{i,j-1}}{1 - k_i^2} \quad (j=1, 2, \dots, i-1) \dots\dots\dots (11)$$



(a) 音響管モデル



(b) 推定法の流れ図

図 1 声道断面積関数の推定

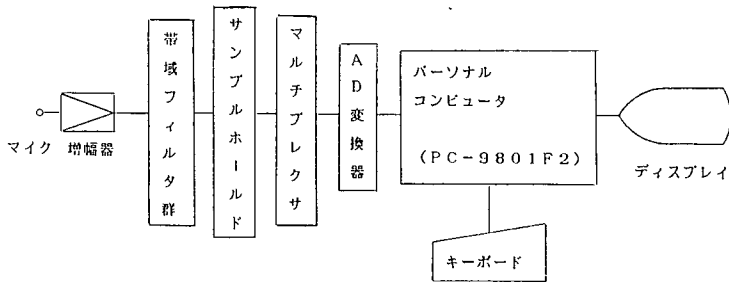


図2 システム構成

$$k_i = a_{i,i}$$

ここで i は p から $p-1, \dots, 1$ と減少してゆき、また最初は

$$a_{p,i} = a_i (i=1, 2, \dots, p)$$

とする。

以上の計算過程を図示すれば、図1(b) のようであり、音声波形のスペクトルから声道断面積関数を推定できる。

2.2 システム構成

図2に、音声波形から声道断面積関数を推定するシステムの構成図を示す。本システムは、音声波形の周波数スペクトルを求める周波数分析部と、このスペクトルから声道断面積関数を計算し、声道モデルをディスプレイ上に表示するパーソナルコンピュータ部(PC-9801)とから構成されている。

まず、周波数分析部である帯域フィルタ群によって、音声波形を約 10[msec] ごとにスペクトル分析し、サンプルホールド回路、マルチプレクサ及び AD 変換器を介して、パーソナルコンピュータに入力する。次に、パーソナルコンピュータは、(6)式に示す対数スペクトルを求め、次の離散的フーリエ逆変換から $C(n)$ を求める。

$$C(n) = \frac{1}{N} \sum_{k=0}^{N-1} (-\log|S(k)|) e^{j\frac{2\pi}{N}kn} \dots\dots\dots(12)$$

ただし、

$$\log|S(k)| = \log|S(N-k)|, \quad k=1, 2, \dots, p$$

本研究では帯域フィルタ群は17チャンネルあり、 $N=32(=2^5)$ 、 $p=16$ としてある。そこで、上式の計算は FFT (高速フーリエ変換) プログラムによって行う。

更に、(8)~(11)式を用いて声道の断面積関数を計算する。計算結果は、断面図としてディスプレイ上に表示する。

プログラムは、スペクトルデータの取り込みにアセンブリ言語を用い、他はすべて FORTRAN で記述している。そのため、現状では1秒間に4~5回程度しか断面積が推定できないが、母音のように比較的安定した発声に対しては十分実用しうると考える。一方、子音については、一旦スペクトルデータをメモリに格納し、その後、計算を実行すれば、声道断面積の推

定が可能である。

3. 実験結果

3.1 推定例

成人男性の発声した5母音（安居院氏ら²⁾のデータによる）に対する声道断面積関数の推定

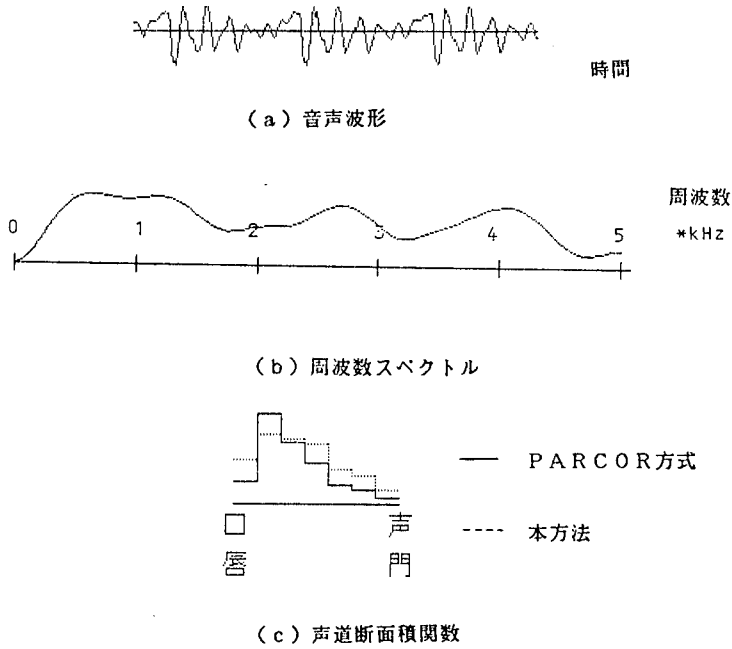


図3 母音/あ/

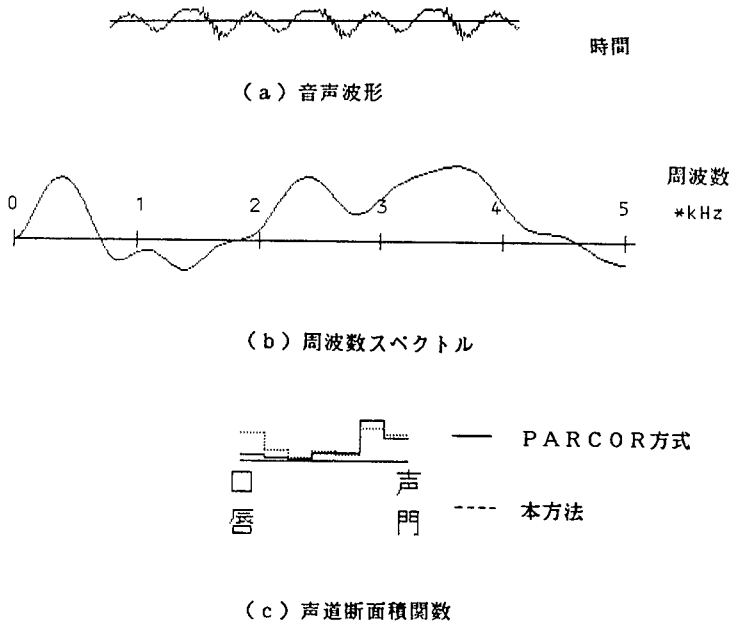
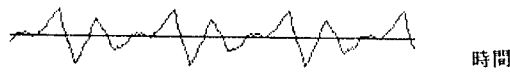
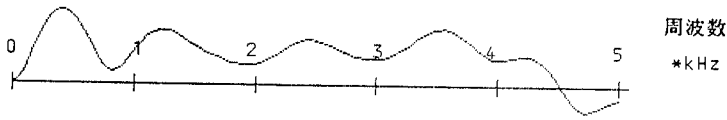


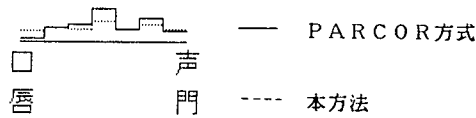
図4 母音/い/



(a) 音声波形

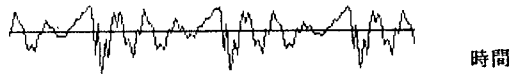


(b) 周波数スペクトル

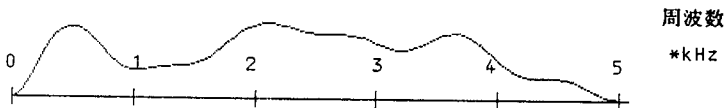


(c) 声道断面積関数

図 5 母音/う/



(a) 音声波形



(b) 周波数スペクトル



(c) 声道断面積関数

図 6 母音/え/

結果の例を図3～図7に示す。図において、(a)は音声波形、(b)は周波数スペクトル、及び(c)は推定された声道断面積関数を表わす。ただし、面積関数は、声門面積 (A_8 , $i=7$ まで図示) が一定となるように正規化してある。なお、周波数スペクトルは、本方法が真にどの程度の精度をもっているかを確認するために、高速フーリエ変換によって得られたスペクトルを用

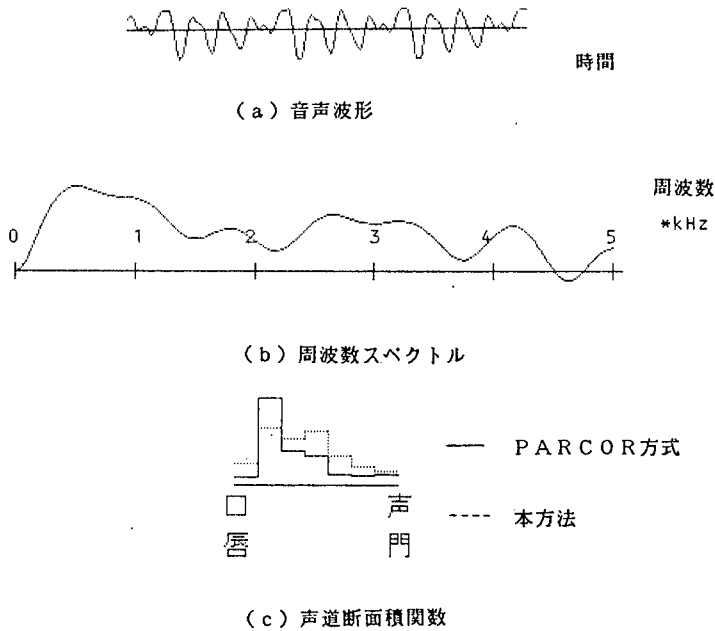


図 7 母音／お／

いている。

図 3 (c)～図 7 (c) において、実線は、現状で最も推定精度が高いとされている PARCOR 方式によるもので、点線が本方法によるものである。図から、／お／において若干精度が劣るものの、比較的類似した声道形状が得られている。

なお、子音については、現在検討中であるので、詳細は別稿としたい。

3.2 本方法の評価

声道断面面積推定における主要な問題は、声道の共振周波数（ホルマント周波数）と帯域幅に影響している因子の影響を、いかに音声波形あるいはスペクトルから除去するかにある。しかしながら、これらの因子が、具体的に声道形状にどのような影響を与えるかは、あまり研究されていないようである。そこで、本節では、正確なスペクトルが求まったと仮定して、これらの因子による影響と、本方法の推定精度について検討する。

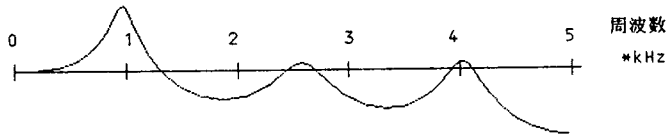
現状で最も正確に音声波形のスペクトルを求める方法として、線形予測分析法がある。この方法は、まず音声波形から直接に自己相関法あるいは共分散法により、予測係数 $\{a_i\}$ を求める。予測係数と声道特性とは(4)式の関係がある。そこで、(4)式の対数をとると

$$\log|H(e^{j\omega})| = -\log|1 + \sum a_i e^{-ij\omega}|$$

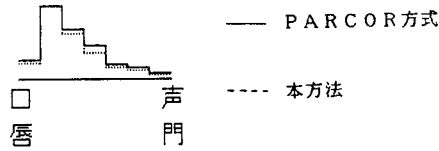
となり、これは、ピッチ周波数に影響されない滑らかなスペクトル包絡特性を表わす。右辺において

$$1 + \sum a_i \exp(-ij\omega)$$

の計算は FFT によって容易に実行できる。このようにして得られたスペクトルと、このスペクトルからの声道形状の推定結果を図 8～図 12 に示す。図において声道形状は最大開口部面積

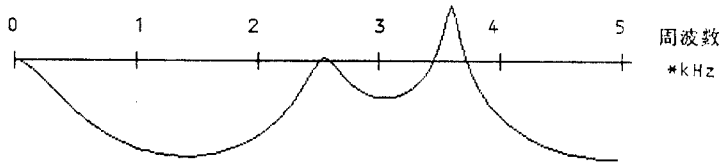


(a) 周波数スペクトル

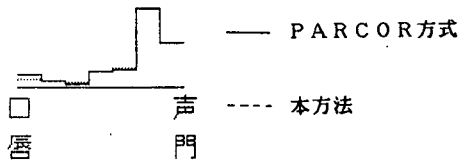


(b) 声道断面積関数

図 8 母音/あ/



(a) 周波数スペクトル

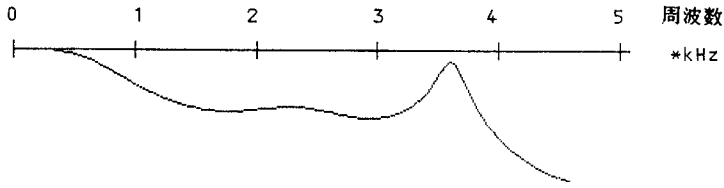


(b) 声道断面積関数

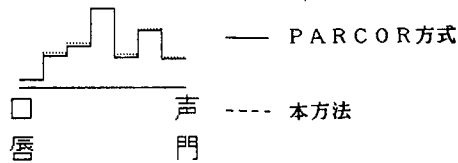
図 9 母音/い/

が一定となるように正規化してある。図から明らかなように、正しいスペクトルが与えられた場合の推定精度は極めて高い。又、FFT によって得られたスペクトル（図3～図7）と比べてみると、線形予測法によるスペクトルは、ホルマント近傍のピークが顕著に現われており、音源特性の影響が少ないことがわかる。この音源の影響は声道の推定結果からみると、全体の形状をなだらかにする傾向があり、特に口唇部の開口面積に大きな影響を与えている。しかし、調音位置（調音に際して閉鎖あるいは狭まりの起こる場所）については、比較的影響が少ない。

以上のように、本方法は、音源特性の影響が少ないスペクトルが得られれば、極めて高い推定精度が可能である。しかし、線形予測法に匹敵するスペクトルを実時間で得ることは、現状では極めて困難である。そこで、筆者らは、この点を改善すべく、信号処理専用のプロセッサ

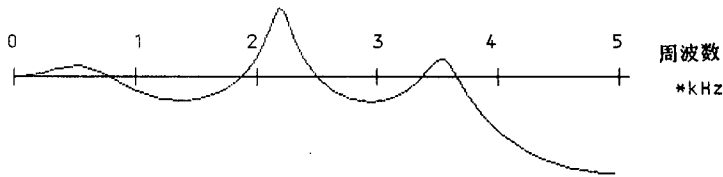


(a) 周波数スペクトル

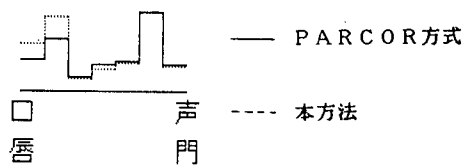


(b) 声道断面積関数

図10 母音/う/



(a) 周波数スペクトル



(b) 声道断面積関数

図11 母音/え/

を用いた分析装置の試作を進めている。

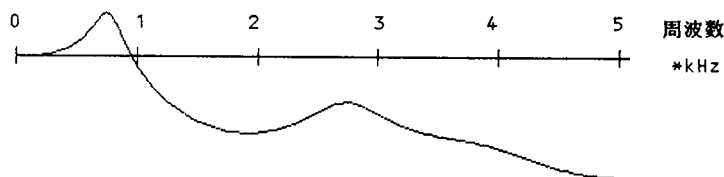
4. む す び

複素ケプストラムを用いた声道断面積関数の推定法について述べた。

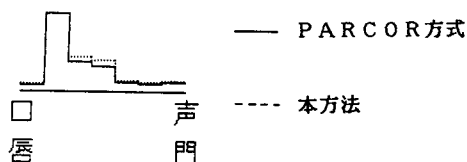
本研究によって得られた結果は、次のとおりである。

- (1) 計算が比較的複雑で時間の要する予測係数の抽出を、ハードウェアとFFT演算によって行なっているので、高速推定が可能である。
- (2) 音源特性の影響の少ないスペクトルを与えれば、高精度の推定結果を得ることができる。

ケプストラムに基づく声道断面積関数の推定



(a) 周波数スペクトル



(b) 声道断面積関数

図12 母音/お/

今後の課題としては、より正確なスペクトルを実時間で得ることと、計算の高速化である。

終りに、日頃ご指導いただいている本学工学部本多高教授に感謝の意を表します。又、本研究にご協力いただいた本学計測制御研究室、電子回路研究室および電子機器研究室の学生諸氏に識意を表する。なお、資材面でご協力いただいたアジアエレクトロニクス㈱、焼結金属工業㈱および高木産業㈱に感謝します。更に、本研究で用いている音声サンプルは、東京工業大学教授安居院博士らが採集したものであり、記して深謝致します。

参 考 文 献

- 1) 石田, 鎌田, 小川: 汎用マイクロコンピュータによる発話訓練機の試作, 明大科研紀要, Vol. 23, No. 8 (1984).
- 2) 安居院, 中嶋: コンピュータ音声処理, 産報出版, pp. 43-81 (1980).
- 3) 伊達訳: デジタル信号処理 (下), コロナ社, pp. 168-180 (1978).
- 4) 鈴木訳: 音声の線形予測, コロナ社, pp. 85-117 (1980).